

Amendments to the Claims:

New claim 102 has been added. This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Previously Presented) A method of providing load balancing among host servers in a computer network using a load balancing switch and a plurality of site switches, the method comprising:

collecting at said load balancing switch a first set of performance metrics regarding said computer network, each said host server being accessed through at least one said site switch;

receiving, at said load balancing switch, a plurality of network addresses generated by an authoritative domain name system server in response to a query regarding a domain name, the authoritative domain name system server and the load balancing switch being separate network devices;

arranging, at said load balancing switch, said network addresses as an ordered list in accordance with at least some of said first set of performance metrics; and

forwarding said ordered list of network addresses as a response to said query.

2. (Previously Presented) A method as in Claim 1, further comprising:

collecting a second set of performance metrics regarding said computer network at each of said site switches;

sending said second set of performance metrics from said site switches to said load balancing switch; and

arranging, at said load balancing switch, said network addresses as an ordered list in accordance with at least some of said second set of performance metrics.

3. (Original) The method of Claim 1, wherein said first set of performance metrics includes a health check sent from said load balancing switch to each of said site switches.

4. (Previously Presented) The method of Claim 3 wherein, when a particular said site switch fails said health check, a network address of said particular site switch is provided a lesser position in said ordered list or is dropped.

5. (Previously Presented) The method of Claim 2, wherein said second set of performance metrics includes a number of sessions connected to host servers having network addresses configured on the respective site switch.

6. (Previously Presented) The method of Claim 5, wherein when said number of sessions at a particular said site switch exceeds a threshold percentage of said particular site switch's maximum session capacity, a corresponding one of said network addresses is provided a lesser position in said ordered list.

7. (Previously Presented) The method of Claim 2, wherein said collecting said second set of performance metrics includes recording, at each said site switch, a round trip time indicative of elapse time for exchanging messages between the respective site switch and a client machine of said computer network.

8. (Previously Presented) The method of Claim 7, wherein said round trip time is an actual recorded time period between the respective site switch receiving a connection request from said client machine and the respective site switch receiving an acknowledgement of a connection from said client machine.

9. (Previously Presented) The method of Claim 1, wherein said arranging takes into consideration a geographical location of an originator of said query.

10. (Previously Presented) The method of Claim 3, wherein said collecting of said first set of performance metrics includes recording a time interval for each said site switch

between said load balancing switch initiating said health check and said load balancing switch receiving a response from said site switch.

11. (Original) The method of Claim 1, wherein said arranging selects a network address of a least recently selected host server for placement at a higher position in said ordered list.

12. (Original) The method of Claim 1 further comprising said load balancing switch limiting a valid time for each network address in said ordered list to less than a predetermined value.

13. (Previously Presented) The method of Claim 1, further comprising, when a connection request is received at a particular said site switch for a connection to one of said host servers, said particular site switch redirecting said connection request to another one of said host servers.

14. (Previously Presented) A system for balancing load among host servers of a computer network, comprising:

an authoritative domain name system server;
a load balancing switch coupled to said authoritative domain name system server, said load balancing switch (a) being a separate network device from said authoritative domain name system server; (b) capable of collecting a first set of performance metrics regarding said computer network; and (c) capable of arranging a list of network addresses received from said authoritative domain name system server in accordance with said first set of performance metrics, the list of network devices being generated by the authoritative domain name system server in response to a query regarding a domain name; and

a plurality of site switches coupling said host servers to said computer network.

15. (Previously Presented) A system as in Claim 14, wherein each of said site switches (a) collects a second set of performance metrics regarding said computer network, and (b) sends said second set of performance metrics to said load balancing switch, said load

balancing switch being capable of using the received second set of performance metrics to arrange the list of network addresses.

16. (Original) A system as in Claim 14, wherein said first set of performance metrics includes a health check sent from said load balancing switch to each of said site switches.

17. (Previously Presented) A system as in Claim 16, wherein, when a particular said site switch fails said health check, a network address of said particular site switch is provided a lesser position in said ordered list.

18. (Previously Amended) A system as in Claim 15, wherein said second set of performance metrics includes a number of sessions connected to a network address configured at said respective site switch.

19. (Previously Presented) A system as in Claim 18, wherein when said number of sessions exceeds a threshold percentage of said respective site switch's maximum capacity, said network address can be provided a lesser position in said ordered list.

20. (Previously Presented) A system as in Claim 15, wherein said second set of performance metrics includes a round trip time indicative of elapse time for exchanging messages between said respective site switch and a client machine of said computer network.

21. (Previously Amended) A system as in Claim 20, wherein said round trip time is an actual recorded time period between said respective site switch receiving a connection request from said client machine and said respective site switch receiving an acknowledgement of a connection from said client machine.

22. (Previously Presented) A system as in Claim 14, wherein said first set of performance metrics includes a geographical location of an originator of said query.

23. (Previously Presented) A system as in Claim 17, wherein said first set of performance metrics includes a time interval for each said site switch between said load balancing switch initiating said health check and said load balancing switch receiving a response from the respective site switch.

24. (Original) A system as in Claim 14, wherein said arranging selects a network address of a least recently selected host server for placement at a higher position in said ordered list.

25. (Original) A system as in Claim 14, wherein said load balancing switch limits a valid time for each network address in said ordered list to less than a predetermined value.

26. (Previously Presented) A system as in Claim 14, wherein when a connection request is received at a said site switch for a connection to one of said host servers, said site switch redirects said connection request to another one of said host servers.

27-69. (Cancelled).

70. (Previously Presented) A method of load balancing among host servers of a data network, the method comprising:

storing, in a load balancing switch of the data network, round trip time data, wherein the round trip time data is a time for exchanging at least one message between a first host server site switch of the data network and a first client machine of the data network; and

ordering, in the load balancing switch, a plurality of network addresses, the network addresses being responsive to a query regarding a domain name, wherein the load balancing switch is capable of ordering the plurality of network addresses based, at least in part, on the round trip time data.

71. (Previously Presented) The method of claim 70, further comprising: creating a table, in the load balancing switch, using the round trip time data.

72. (Previously Presented) The method of claim 71, wherein the table is indexed by network neighborhood.

73. (Previously Presented) The method of claim 70, further comprising: sending a health check message to each of the plurality of network addresses from the load balancing switch.

74. (Previously Presented) The method of claim 73, wherein the health check is a layer 7 health check.

75. (Previously Presented) The method of claim 74, wherein the layer 7 health check uses a HTTP protocol or a FTP protocol.

76. (Previously Presented) The method of claim 73, wherein the health check is a layer 4 health check.

77. (Previously Presented) The method of claim 76, wherein the health check is a TCP or UDP health check.

78. (Previously Presented) The method of claim 70, wherein the first host server site switch is one of a plurality of host server site switches of the data network, and the first client machine is one of a plurality of client machines of the data network, and further comprising:

storing, in the load balancing switch, round trip time data received from each of the plurality of host server site switches, wherein each said round trip time data is a time for exchanging at least one message between a respective one of the host server site switches and a respective one of the plurality of client machines.

79. (Previously Presented) The method of claim 78, further comprising: aggregating, at the load balancing switch, the round trip time data received from the plurality of host server site switches into a table.

80. (Previously Presented) The method of claim 79, wherein the table is indexed by network neighborhood.

81. (Previously Presented) The method of claim 78, wherein the time for exchanging messages between the respective host server site switch and the respective client machine is a time difference between the receipt, at the respective host server site switch, of a connection request message and a connection acknowledgement message from the respective client machine.

82. (Previously Presented) The method of claim 81, wherein the connection request message and the connection acknowledgment message comprise a TCP SYN packet and an associated TCP ACK packet, respectively.

83. (Previously Presented) the method of claim 78, wherein the query originated at the first client machine; and further comprising:

ordering, in the load balancing switch, the network addresses from a first said network address to a last said network address, wherein the first said network address is associated with the one of the plurality of host server site switches having a lowest round trip time with the first client machine.

84. (Previously Presented) The method of claim 83, wherein among the ordered network addresses a second said network address is associated with the one of the plurality of host server site switches having a next-to-lowest round trip time with the first client machine.

85. (Previously Presented) The method of claim 70, wherein the time for exchanging messages between the first host server site switch and the first client machine is a time difference between the receipt, at the first host server site switch, of a connection request message and a connection acknowledgement message from the first client machine.

86. (Previously Presented) A method of load balancing among host servers of a data network, the method comprising:

receiving, at a load balancing switch of the data network, a query regarding a domain name; and

selecting, from a plurality of network addresses responsive to the request, a best network address based, at least in part, on which of the plurality of network addresses has been least recently selected by the load balancing switch as a best network address in response to previous queries.

87. (Previously Presented) The method of claim 86, further comprising:

storing, at the load balancing switch, round trip time data, wherein each said round trip time data is a time for exchanging at least one message between a respective one of a plurality of host server site switches of the data network and a respective one of a plurality of client machines of the data network

88. (Previously Presented) The method of claim 87, wherein the time for exchanging messages between the respective host server site switch and the respective client machine is a time difference between the receipt, at the respective host server site switch, of a connection request message and a connection acknowledgement message from the respective client machine.

89. (Previously Presented) The method of claim 87, further comprising:

creating, in the load balancing switch based on the round trip time data, a proximity table.

90. (Previously Presented) The method of claim 86, further comprising:

creating, in the load balancing switch, a proximity table.

91. (Previously Presented) A load balancing switch for load balancing amongst a plurality of host servers of a data network, the load balancing switch comprising:

a means for storing round trip time data received from a plurality of host server site switches, each said round trip time data being a time for exchanging at least one message between a respective one of said host server site switches and a respective one of a plurality of client machines of the data network;

a means for receiving a query regarding a domain name, the query originating at a client machine of the data network; and

a means for ordering a plurality of network addresses that are responsive to the query based, at least in part, on stored round trip time data for the query-originating client machine.

92. (Previously Presented) The load balancing switch of claim 91, further comprising:

a means for ordering the plurality of network addresses based, at least in part, on which of the network addresses has been least recently selected as a best network address in response to previous queries.

93. (Previously Presented) The load balancing switch of claim 91, further comprising:

a means for ordering the plurality of network addresses based, at least in part, on a session capacity of the plurality of host server site switches.

94. (Previously Presented) The load balancing switch of claim 91, further comprising:

a means for ordering the plurality of network addresses based, at least in part, on an available session capacity of the plurality of host server site switches, wherein the available session capacity is a percentage of a session capacity of the respective host server site switch.

95. (Previously Presented) The load balancing switch of claim 91, further comprising:

a means for ordering the plurality of network addresses based, at least in part, on a health of the host servers.

96. (Previously Presented) A load balancing switch for load balancing amongst a plurality of host servers of a data network, the load balancing switch comprising:
a means for receiving a query regarding a domain name; and
a means for ordering a plurality of network addresses that are responsive to the query, based, at least in part, on which of the network addresses has been least recently selected as a best network address in response to previous queries.

97. (Previously Presented) The load balancing switch of claim 96, further comprising:

a means for ordering the plurality of network addresses based, at least in part, on a session capacity of a plurality of host server site switches, each said host server site switch being coupled between the load balancing switch and at least one of the host servers.

98. (Previously Presented) The load balancing switch of claim 96, further comprising:

a means for ordering the plurality of network addresses based, at least in part, on an available session capacity of a plurality of host server site switches, each said host server site switch being coupled between the load balancing switch and at least one of the host servers, wherein the available session capacity is a percentage of a session capacity of the respective host server site switch.

99. (Previously Presented) The load balancing switch of claim 96, further comprising:

a means for ordering the plurality of network addresses based, at least in part, on a health of the host servers.

100. (Previously Presented) A data networking method comprising:

storing, in a host server site switch through which a plurality of host servers of a data network are accessed, round trip time data, the round trip time data being a time for exchanging at least one message between the host server site switch and a client machine of the data network; and

communicating the round trip time data to a load balancing switch of the data network.

101. (Previously Presented) The data networking method of claim 100, further comprising communicating a number of sessions of the host server site switch to the load balancing switch.

102. (New) A method of providing load balancing among host servers in a computer network using a load balancing switch and a plurality of site switches, each site switch coupled to one or more host servers, the method comprising:

receiving, at the load balancing server from at least one site switch from the plurality of site switches, metrics information regarding one or more host servers coupled to the at least one site switch;

receiving, at the load balancing switch from a domain name system server, a plurality of network addresses generated by the domain name system server in response to a query regarding a domain name, wherein the domain name system server is an authoritative domain name system server for multiple domains; and

arranging, at the load balancing switch, the plurality of network addresses as an ordered list based upon the metrics information; and

forwarding the ordered list of network addresses as a response to said query.